scientific reports

Genome-wide characterization and expression profiling of MADS-box family genes during organ development and drought stress in *Camelina sativa* L

Somayeh Tahmasebi[1], Parisa Jonoubi[1]✉, Mohammad Majdi[2], Ahmad Majd[3] & Parviz Heidari[4]✉

Camelina sativa **stands out among oilseed crops due to its remarkable resilience to challenging environmental conditions and its significant potential for biodiesel production. The MADS-box transcription factors play a pivotal role in numerous biological processes within plants, encompassing growth, development, and responses to environmental stressors. In this research, by employing the BLAST, we have successfully identified 325** ***MADS-box*** **genes within** ***Camelina sativa*** **genome. These genes were systematically categorized into two principal groups: type I (comprising Mα, Mβ, and Mγ) and type II (including MI-KC^C and MIKC*) predicated upon their phylogenetic relationships, structural protein motifs, and exon-intron configurations. Our findings reveal that type II** ***MADS-box*** **genes have, in general, experienced a more profound expansion relative to type I genes. Specifically, the TM3 subgroup within type II** ***MADS-box*** **genes exhibited the highest degree of gene expansion, comprising 21** ***TM3*** **genes. The amino acid sequences encoded by these genes exhibited a length variation ranging from 150 to 820 aa. The predicted molecular weights (MW) of the CsMADS-box proteins displayed a range from 17.01 to 94.06 kDa, while the isoelectric points (pI) were observed to span from 4.13 to 10.09. Evolutionary analysis predicated on the Ka/Ks ratios indicates that the evolutionary pathway of** ***MADS-box*** **genes in** ***Camelina sativa*** **has been predominantly driven by the mechanisms of purifying selection. Moreover, an investigation of cis-acting elements has elucidated the participation of** ***MADS-box*** **genes in the adaptive responses to abiotic stressors. The expression profiles of six Type I and three MIKC^C genes across diverse organs and under varying drought treatment conditions demonstrated that these genes are expressed in both reproductive and vegetative structures, displaying uniform expression patterns throughout several developmental phases of flowering. The expression levels of** ***CsMADS035*****,** ***CsMADS115*****,** ***CsMADS131*****, and** ***CsMADS181*** **were notably modified in reaction to drought stress conditions. The detailed annotation and comprehensive transcriptome profiling provided in this research yield essential insights into the functional roles that** ***MADS-box*** **genes perform in stress resistance, as well as their contributions to growth and developmental processes. This acquired understanding establishes a foundational framework for the functional characterization and potential genetic engineering initiatives pertaining to** ***Camelina sativa*****, thereby augmenting the prospective application of these candidate genes.**

Keywords MADS-box, Expression patern, Abiotic stress, Phylogenetic analysis, Camelina sativa

Plants face a variety of environmental stresses and seasonal fluctuations, such as changes in light, precipitation, nutrient availability, temperature, and UV exposure. These factors can affect numerous metabolic pathways,

[1]Department of Plant Sciences, Faculty of Biological Sciences, Kharazmi University, Tehran, Iran. [2]Agricultural Biotechnology Department, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran. [3]Biology Department, College of Bioscience, Tehran North Branch, Islamic Azad University, Tehran, Iran. [4]Faculty of Agriculture, Shahrood University of Technology, Shahrood, Iran. ✉email: jonoubi@khu.ac.ir; heidarip@shahroodut.ac.ir

resulting in alterations in plant development and morphology. Nevertheless, plants have developed innate mechanisms to adapt and thrive in diverse environments[1].

The development of plants and their responses to dynamic environmental conditions and stressors are regulated by intricate networks that involve multiple molecular components, including transcription factors (TFs), regulatory RNAs, and enzymes[2]. Among these, MADS-domain transcription factors are pivotal elements within the regulatory frameworks that govern various developmental processes in plants, animals, and fungi[3]. The MADS-box gene family is instrumental in modulating plant growth, with particular emphasis on the development of floral organs and the duration of flowering[4]. The progress in whole-genome sequencing has facilitated the comprehensive identification of MADS-box gene families across numerous plant species, including *Setaria italic*, *Chrysanthemum nankingense*, *Brassica oleracea*, *Brassica rapa*, *Brachypodium distachyon*[510]. Beyond growth and development, the MADS-box gene family also influences plant responses to abiotic stresses[1]. For example, in maiz, the gene ZMM7-L is upregulated in response to NaCl, which diminishes seed germination in conditions of salinity stress, thereby suggesting a deleterious regulatory function in the context of salt tolerance[11] Conversely, in tomato, the genes SlMBP8 and SlMBP11 serve to enhance resistance to saline-alkaline environments[12]. *OsMADS26* in rice negatively regulates disease and drought resistance, highlighting its involvement in multiple abiotic stress responses[13]. These investigations underscore the significance of MADS-box genes in enhancing stress resistance, thereby offering crucial resources for the advancement of plant breeding and the improvement of agricultural crops[14]. By utilizing floral homeotic mutants, researchers have been able to elucidate the widely recognized ABCDE model for the specification of floral structures[15]. Numerous investigations into this gene family have revealed that the genes within this group play a crucial role in the development of flowers[16–18].

The MADS-box family encodes a specific class of transcription factors characterized by a conserved region comprising 58–60 amino acids, referred to as the MADS domain[19]. This domain was derived from the initial four identified representatives: MINICHROMOSOME **M**AINTENANCE 1 (MCM1) from *Saccharomyces cerevisiae*[20], **A**GAMOUS (AG) from *Arabidopsis thaliana*[21], **D**EFICIENS (DEF) from *Antirrhinum majus*[22], and **S**ERUM RESPONSE FACTOR (SRF) from *Homo sapiens*[23]. Through an analysis of evolutionary relationships and sequence characterization, Alvarez-Buylla and colleagues categorized MADS-box proteins into two principal types, designated as Type I and Type II, both of which encompass the MADS-box domain[24]. In the plant kingdom, Type I MADS-box genes typically exhibit one or two exons and either none or one intron, encoding proteins characterized by a highly conserved SRF-like MADS domain while being devoid of a K domain[25]. In contrast, Type II MADS-box genes display greater complexity, featuring multiple exons and introns. The proteins they encode have four distinct domains: a conserved MEF2-like MADS (M) domain at the N-terminus, an intervening (I) domain, a semi-conserved keratin-like (K) domain, and a C-terminal (C) domain. These proteins are classified as MIKC*-type MADS-box proteins[26,27]. The M domain demonstrates a significant level of conservation and serves an essential function in nuclear localization, dimerization, DNA binding, and interactions with auxiliary factors[25,28,29]. The I domain aids in protein dimerization[26], while the K domain is essential for forming higher-order complexes and further dimerization[30,31]. The C domain acknowledged for its considerable variability, plays a crucial role in facilitating transcriptional activation[23]. Owing to the inherent variability present within the I and K domains, Type II MADS-box proteins are additionally classified into two distinct categories: MIKC* and MIKC^C. MIKC^C proteins are distinguished by a comparatively shorter I domain and a more conserved K domain when compared to MIKC* proteins[32].

Four Mα subfamilies[33] and two Mγ subfamilies in *Arabidopsis*[34] have identified that these Type I *MADS-box* genes play crucial roles in the process of seed development. In contrast, studies on Type II genes are more extensive, covering 13 subfamilies that are classified under the MIKC^C type[35].

Although MADS-box transcription factors exhibit promise in regulating stress responses, the precise functions of many of these factors are not completely understood. Despite the considerable efforts in conducting genome-wide investigations and functional evaluations of MADS-box proteins across a variety of species, to date, no detailed genome-wide analysis and systematic classification of *MADS-box* genes in *Camelina sativa* have been conducted, highlighting a valuable opportunity for in-depth investigation and discovery. *Camelina sativa*, commonly referred to as camelina, is an annual C4 plant native to southeastern Europe and southwestern Asia, thriving in temperate climates[36]. It is a member of the Crucifer family (Brassicaceae) and shares a close evolutionary relationship with *Arabidopsis thaliana*[37].

Camelina is a versatile oilseed crop with a wide array of applications extending beyond biofuel production. Camelina oil, known for its high content of omega-3 fatty acids and a well-balanced profile of omega fatty acids, particularly alpha-linolenic acid (ALA), holds significant value for human nutrition. Its benefits for cardiovascular health have made it an increasingly popular ingredient in dietary supplements[38]. The oil also finds use in the cosmetics and personal care industry for its moisturizing properties, while the by-product, camelina meal, serves as a nutritious protein source for livestock and poultry feeds[39]. Additionally, camelina demonstrates its potential as a sustainable biobased crop for marginal lands and proves its adaptability across challenging environments[40].

Overall, *Camelina sativa* presents diverse opportunities across the food, health, sustainable agriculture, and materials science sectors. As a major vegetable crop globally, camelina faces numerous abiotic stresses. Besides, the identification of MADS-box genes associated with stress resistance in *Camelina sativa* may yield significant advantages. So far, a comprehensive study on the identification and function of MADS-box gene family in Camelina has not been conducted. In the present investigation, we analyzed the genomic distribution, structural motifs of proteins, phylogenetic affiliations, and gene architectures of all identified *MADS-box* candidates within *C. sativa*. Furthermore, we performed expression patterning for nine *MADS-box* genes, encompassing both Type I and Type II subfamilies, across four vegetative organs, five developmental stages of flowering, three phases of seed maturation, and in response to drought stress.

Materials and methods

Identification of *MADS-box* genes

The amino acid sequences of MADS-box genes derived from *Arabidopsis thaliana* were extracted from the Arabidopsis Information Resource (TAIR10) database (https://www.arabidopsis.org). These sequences were subsequently utilized as queries in the Basic Local Alignment Search Tool (BLAST) to ascertain the presence of additional MADS-box gene homologs within *Camelina sativa*, employing the Ensembl Plants database (https://plants.ensembl.org/index.html). To corroborate the presence of the MADS-box domain in the identified proteins, we utilized the Conserved Domain Database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml).

Biochemical characteristics analysis

The molecular weight (MW), isoelectric point (pI), aliphatic index, instability index, and grand average of hydropathy (GRAVY) values were estimated using the ProtParam utility, which is accessible on the ExPASy database (http://web.expasy.org/protparam).

Phylogenetic analysis

The amino acid sequences of MADS-box proteins from selected plants were aligned utilizing the Clustal Omega tool[41]; https://www.ebi.ac.uk/Tools/msa/clustalo). Following the alignment, a phylogenetic tree was generated using the Maximum Likelihood method, supported by 1,000 bootstrap replicates to ensure reliability, employing the IQ-TREE software for the analysis[42]. The resulting phylogenetic tree was visualized and annotated using the iTOL platform[43] (https://itol.embl.de/).

Predicted protein motifs and gene structure characterization of *MADS-box* genes

For the characterization of predicted protein motifs and gene structure of the *MADS-box* genes, the Multiple EM for Motif Elicitation (MEME) online resource was utilized to identify motifs within the anticipated MADS-box proteins[44]. The MEME analysis was conducted under defined parameteres aimed to detect 20 distinct motifs, with widths ranging from 6 to 200 amino acids, while permitting an unrestricted number of repetitions[45]. Furthermore, coding sequences (CDS) and genomic DNA (gDNA) sequences for the *MADS-box* genes were retrieved from the Ensembl Plants database to facilitate the prediction of their gene structures. Subsequently, the gene structures were illustrated using the Gene Structure Display Server 2. (GSDS 2.)[46].

Chromosomal localization and gene duplication analysis

The mapping of 325 *MADS-box* genes identified in the reference genome of *Camelina sativa* to its chromosomes was performed using TBtools software[47]. An analysis of gene duplication was performed with MCScanX[48]. Employing key criteria such as: (a) alignable sequence length covering at least 75% of the longer gene, and (b) similarity of aligned regions being at least 75% [49]. The protein sequences of *Camelina sativa* were subjected to a self-comparison using the BLASTP algorithm, employing a tabular output format (-m 8) and an e-value threshold of less than $1e^{-10}$ and identity > 85% [5,50]. The output from this BLASTP analysis, in conjunction with a streamlined gene location file for *Camelina sativa* that detailed chromosome assignments, gene symbols, and both start and end coordinates, was then processed using MCScanX to identify various types of gene duplications, utilizing the software's standard settings[51]. Additionally, the calculation of non-synonymous (Ka) and synonymous (Ks) substitution rates was conducted with the KaKs_Calculator available at http://code.google.com/p/kaks-calculator/wiki/KaKs_Calculator[52,53].

Cis-regulatory element analysis and expression analysis of MADS-box genes

In terms of analyzing *cis*-regulatory elements and expression profile of the *MADS-box* genes, the promoter regions were investigated by retrieving the 150 bp upstream sequences of each *MADS-box* gene from the Ensembl Plants database (https://plants.ensembl.org/index.html). These upstream sequences were then examined using the PlantCARE tool (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to identify *cis*-regulatory elements (CREs). All identified CREs were subsequently counted and classified accordingly[54].

Plant materials and treatments

The experimental procedure was executed within a controlled greenhouse environment characterized by a diurnal temperature regime of 25/20°C (day/night), a photoperiod of 16 h of light, and 8 h of darkness, along with a relative humidity level of 60%. Seeds of *Camelina sativa* "Soheil cultivar" were supplied from Bisetoon Shafa Co., Kermanshah, Iran. Plants were grown in the greenhouse and the experiment was designed using a randomized complete block design (RCBD) with three biological replications to ensure statistical reliability and minimize experimental variability. For the organ study, fresh samples of roots, stems, leaves, shoots apex, flower buds (in four diameter sizes: < 2 mm (Bud-X2), ~ 2 mm (Bud-2), ~ 4 mm (Bud-4) and ~ 8 mm (Bud-8)), mature flowers and seeds (in three different stages: early seed development (ES), mid seed development (MS) and late seed development (LS)) were harvested and promptly frozen in liquid nitrogen and stored at -80 °C till RNA extraction.

For treatments, at the budding stage (21 days after planting), water-deficit stresses were assessed in three levels of 100% (DS1) 60% FC (DS2), and 30%FC (DS3) over six time periods (0, 4, 8, 12, 24, and 48 h). In each treatment, treated leaves were colected and frozen immediately in liquid nitrogen and stored at -80 °C till RNA extraction.

Total RNA extraction and cDNA synthesis

Total RNA was extracted from various organs using the RNA extraction kit, RNX^{TM}-Plus (a Guanidine/phenol solution; *SinaClon BioScience*, Tehran, Iran), according to the operating manual. The genomic contamination was removed by treating the isolated RNA samples with an RNase-Free DNase I Kit (*SinaClon BioScience*, Tehran, Iran). RNA concentrations (ng/µl) and purity ratios (260/280 nm and 260/230 nm) were determined using a NanoDrop 2000 spectrophotometer (Bio Tek, USA). Total RNA (1 µg) from each sample was used for cDNA synthesis using the Easy™ cDNA Synthesis Kit (Parstous, Mashhad, Iran) following the manufacturer's instructions. The cycling program was set to 42 °C for 30 min in a single cycle.

Quantitative Real-Time PCR (qRT-PCR) analysis

All qRT-PCR analyses were conducted using a LightCycler 96 system (Roche, Germany) with the SYBR Green I Master Mix kit (Bonyakhte, Tehran, Iran) serving as the fluorescent reporter. Negative controls were included by omitting reverse transcriptase to confirm the absence of DNA contamination in the RNA samples. Each PCR reaction was performed in duplicate (Technical replicates) for every sample in a final volume of 13 µl including 0.5 µl of forward/reverse primer (10 µM), 6.5 µl of SYBR green master mix, 1 µl of cDNA, and 4.5 µl of nuclease-free H_2O in 0.1mL PCR strips. The following thermal cycling program was applied: initial polymerase activation at 95 °C for 2 min, followed by a two-temperature cycling process consisting of denaturation at 95 °C for 5 s and annealing-extension at 55 °C for 30 s, repeated for a total of 40 cycles. The program was completed with a melting curve step: 95 °C for 10 s, 58 °C for 30 s, and 97 °C for 1 s. Fluorescence was measured after the final step of each cycle to quantify the PCR products. The $2^{-\Delta Ct}$ method was used to calculate the mRNA fold change of *CsMADS-box* genes in various organs and the $2^{-\Delta\Delta Ct}$ method was used to calculate the relative gene expression of *CsMADS-box* genes in response to drought stress[55]. The primers of *MADS-box* genes were designed using a gene runner and oligo analyzer (Table S1). In this research, *SEC3A* served as the reference gene[55]. In *Arabidopsis*, the SEC3A protein is an essential component of the exocyst complex, which is key to regulating polarized cell growth. Moreover, *Sect. 3* plays a crucial role in controlling polar exocytosis during pollen tube growth[56,57].

Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics 27, employing a two-way analysis of variance (ANOVA) followed by Duncan's post-hoc test. The error bars in the results represent the standard errors of the fold changes in relative gene expression, calculated from three independent biological replicates and duplicate PCR reactions for each sample.

Results

Identification and characterization of *MADS-box* genes

In the *C. sativa* genome, a total of 325 candidate genes that encode the MADS-box domain were discovered. These genes were designated as *CsMADS001* to *CsMADS325* based on where they are located on the chromosomes. The predicted open reading frames (ORFs) of the *CsMADS-box* genes range in length from 453 to 2463 base pairs, encoding amino acid sequences that vary between 150 and 820 amino acids. The estimated molecular weights (MW) of the CsMADS-box proteins range from 17.01 to 94.06 kilo Daltons, while their predicted isoelectric points (pI) vary from 4.13 to 10.09 (Table 1; Table S2).

Phylogenetic analysis of *MADS-box* genes

A phylogenetic dendrogram was constructed utilizing the complete sequences of MADS-box proteins derived from *C. sativa* and *A. thaliana*. Among the 325 identified *CsMADS* genes, these were categorized into two distinct clades: type I (219) and type II (106). Independent phylogenetic trees were generated for each clade, incorporating *MADS-box* genes from both taxa. The type I *CsMADS* genes were further stratified into three distinct subcategories: Mα, Mβ, and Mγ (Fig. 1). A significant majority of type I *MADS-box* genes from each organism congregated, suggesting a close phylogenetic affinity. The Mα subgroup represented the largest fraction, comprising 109 genes, whereas Mβ and Mγ exhibited relatively comparable sizes, with 51 and 59 genes, respectively. Among the type II *CsMADS* genes, there were 86 MIKC^{C}-type and 20 MIKC*-type genes (Fig. 2). Based on known *A. thaliana MADS-box* gene groups, 10 MIKC^{C} evolutionary branches were identified. The TM3-like branch contained the most *CsMADS* MIKC^{C}-type genes, totaling 21. Two subgroups, *AGL16-like* and *AGL12-like* genes, consisted of only three members each. In contrast to *A. thaliana*, the gene clusters associated with AP1-like and *AGL12-like* in *C. sativa* experienced substantial expansion attributable to gene duplication, whereas the *PI* and *AP3-like* gene families lacked identifiable paralogs.

Gene Name	Transcript ID	Chr	CDS (bp)	lengh (A.A)	molecular weight (kDa)	pI	No. of Exon	Group
CsMADS035	**Csa03g001710**	**3**	**939**	**235**	**26519.17**	**9.63**	**2**	**Mα**
CsMADS066	**Csa05g003630**	**5**	**1238**	**213**	**24428.89**	**9.17**	**8**	**TM3-like**
CsMADS098	**Csa07g017080**	7	1939	551	59862.79	4.94	1	Mß
CsMADS115	**Csa08g015010**	8	699	232	26785.58	9.18	1	Mß
CsMADS131	**Csa09g008560**	9	465	154	17318.43	5.24	2	Mα
CsMADS176	**Csa11g088930**	11	813	270	30876.67	5.53	1	Mγ
CsMADS181	**Csa11g102240**	11	1175	195	21985.35	7.77	7	FLC-like
CsMADS265	**Csa17g029500**	17	1712	163	18752.98	9.38	1	Mγ
CsMADS291	**Csa18g034280**	18	1312	242	27649.5	9.42	8	AP1-like

Table 1. Key MADS-box genes in *Camelina sativa* L.

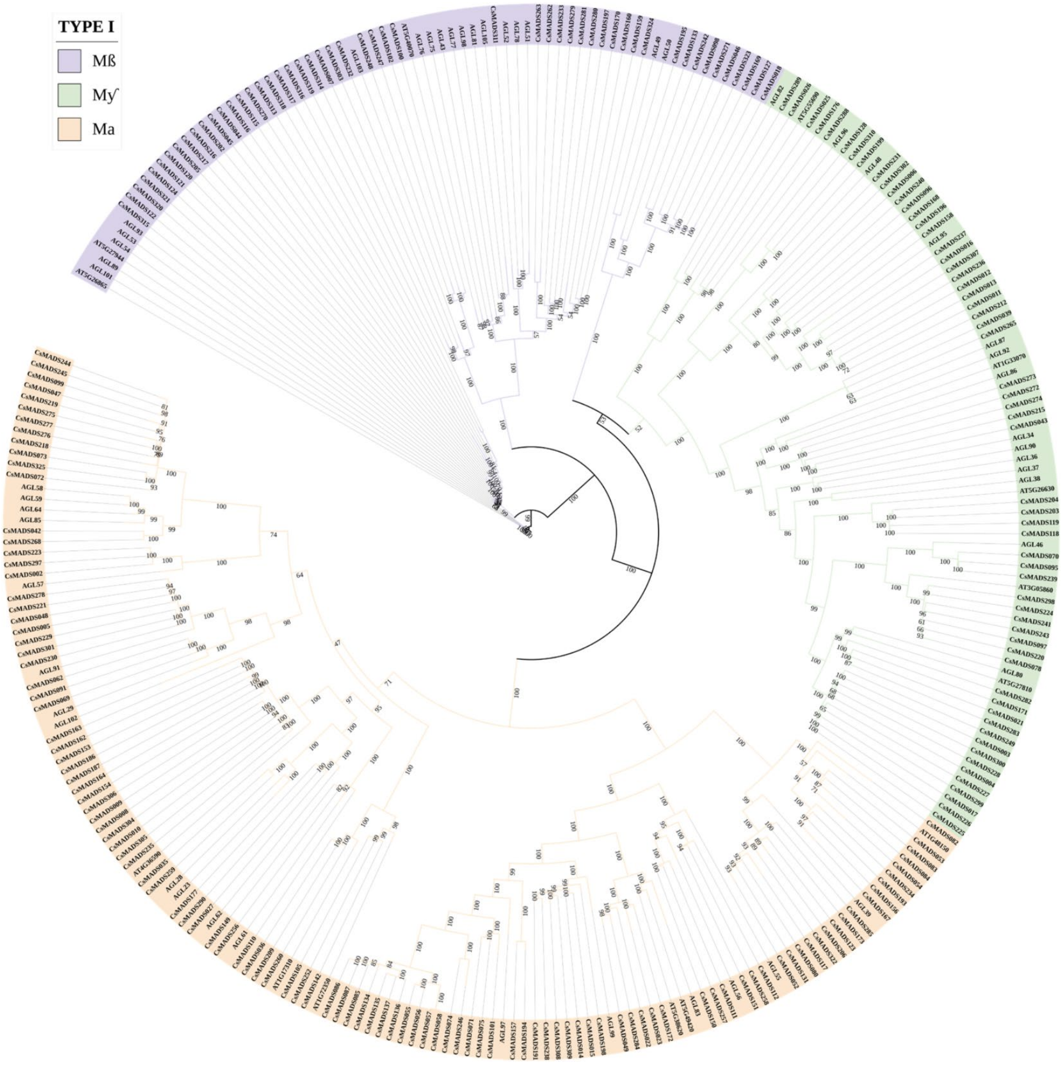


Fig. 1. Phylogenetic analysis of TYPE I-MADS-box transcription factor genes in *Camelina sativa* and *Arabidopsis thaliana.* A total of 283 TYPE I- MADS-box protein sequences were used to construct the ML (maximum likelihood) tree.

An additional phylogenetic dendrogram was established utilizing the full-length MADS-box protein sequences from *C. sativa* (Supplementary Fig. 1). Generally, type II *MADS-box* genes displayed a greater extent of expansion compared to their type I counterparts. Within the type II *MADS-box* gene classification, the TM3 subgroup exhibited the most pronounced degree of expansion, encompassing a total of 21 *TM3* genes.

Conserved motifs analysis of MADS-box proteins

The MEME online tool was employed to delineate conserved motifs among the 325 predicted MADS-box proteins of *C. sativa*. A cumulative total of 10 conserved motifs, designated as motif 1 through motif 10, were discerned (Figs. 3a and b, 4a and b and 5a and b). The Figs. 3-a and b, 4-a and b and 5a and b illustrates that Type I and Type II MADS-box proteins encompass nine and seven principal motifs, respectively. Type I MADS-box proteins demonstrate a greater degree of motif variability, likely attributable to the non-conserved regions at their C-terminal ends. Certain motifs are distinctly characteristic of each family; for instance, motifs 7, 8, and 10 are exclusively associated with Type I proteins, whereas motif 3 is singular to Type II proteins. The specificity of these protein motifs is likely a contributing factor to the functional heterogeneity observed within the Type I and Type II protein families. Some members, including *CsMADS059*, *CsMADS138*, and *CsMADS161*, exhibit distinctions from other constituents of the same subclade with respect to specific motifs. The comprehensive amino acid sequences of the 20 motifs are displayed in Table 2.

Gene structure analysis of *MADS-box* family

To explore gene structures, we analyzed the intron-exon organization of *C. sativa MADS-box* genes utilizing the GSDS online tool. As depicted in Figs. 3-a and c, 4-a and c and 5a and c, we observed variation in the number of introns among these genes, ranging from 0 to 12. Type I and Type II protein genes manifested significant

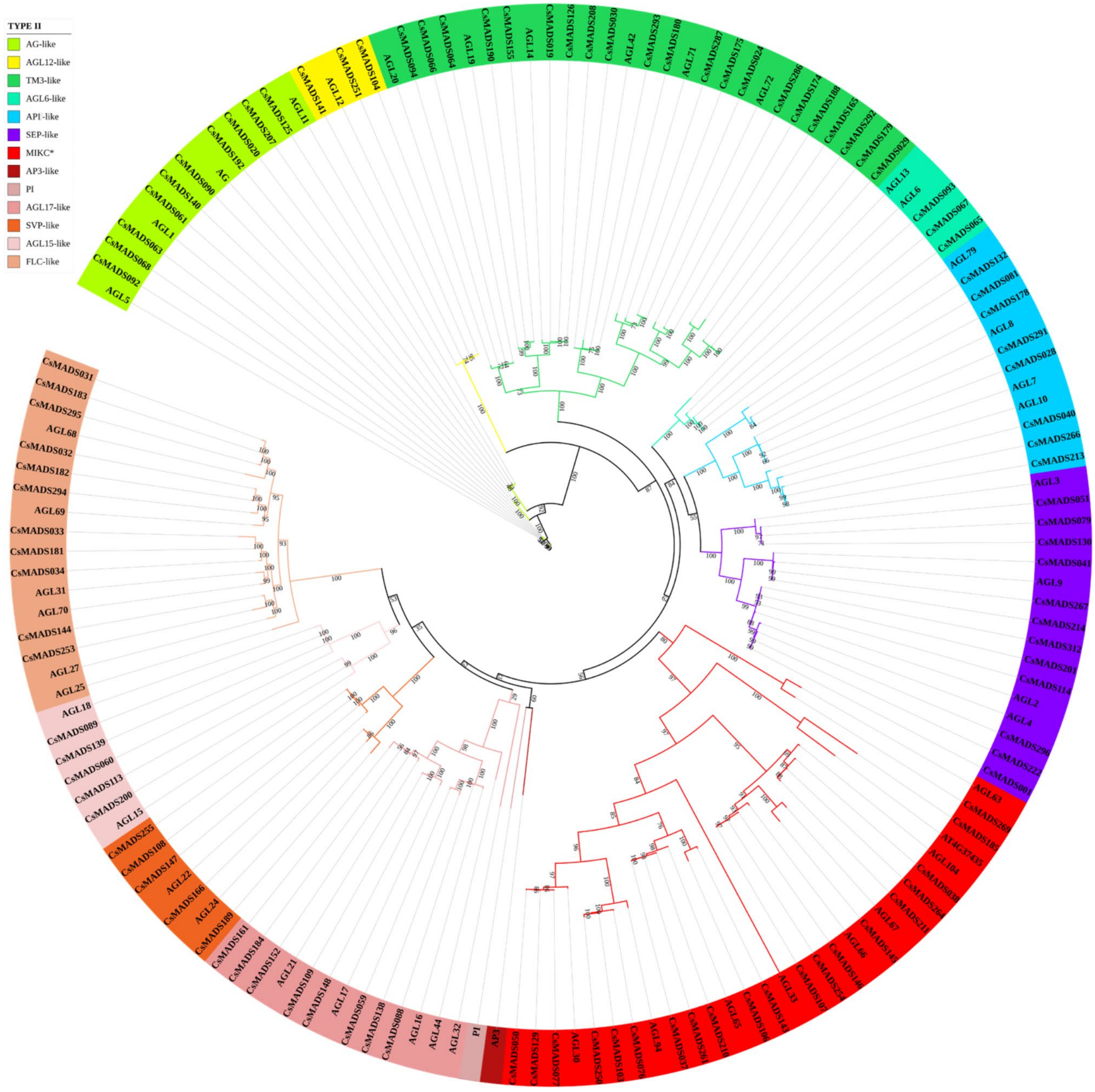


Fig. 2. Phylogenetic analysis of TYPE II-MADS-box transcription factor genes in *Camelina sativa* and *Arabidopsis thaliana*. A total of 153 TYPE II- MADS-box protein sequences were used to construct the ML (maximum likelihood) tree.

discrepancies in the number of introns. Specifically, 20 out of 219 (9.13%) Type I genes contained two or more introns, while all 106 (100%) Type II genes had at least three introns (Fig. 6a). Besides, high variation was observed between subfamilies of type II in term of predicted pI value (Fig. 6b).

Chromosomal locations analysis of *MADS-box* genes

The localization of *MADS-box* genes on the chromosomes of *C. sativa* was elucidated utilizing TBtools software. As illustrated in Fig. 7a, these genes exhibit a distribution across 20 distinct chromosomes. Among the total of 325 genes, 323 were found to be randomly allocated across these chromosomes, whereas 3 genes were situated on 3 scaffolds. Chromosome 10 harbors the least number of *MADS-box* genes, accounting for 2.77%, in contrast to chromosomes 9, 17, and 11, which possess the most significant quantities of *MADS-box* genes (7.08%, 7.08%, and 7.07%, respectively) (Fig. 7b). Moreover, an asymmetrical distribution of Type I and Type II MADS-box genes was discerned on the chromosomes of *C. sativa*. Specifically, a solitary Type I gene was identified on chromosomes 1, 15, 19, and 20, in contrast to the Type II genes, which are dispersed across all 20 chromosomes (Fig. 7). The quantities of Type I and Type II genes on chromosomes 4, 6, and 18 are equivalent, whereas chromosomes 2 and

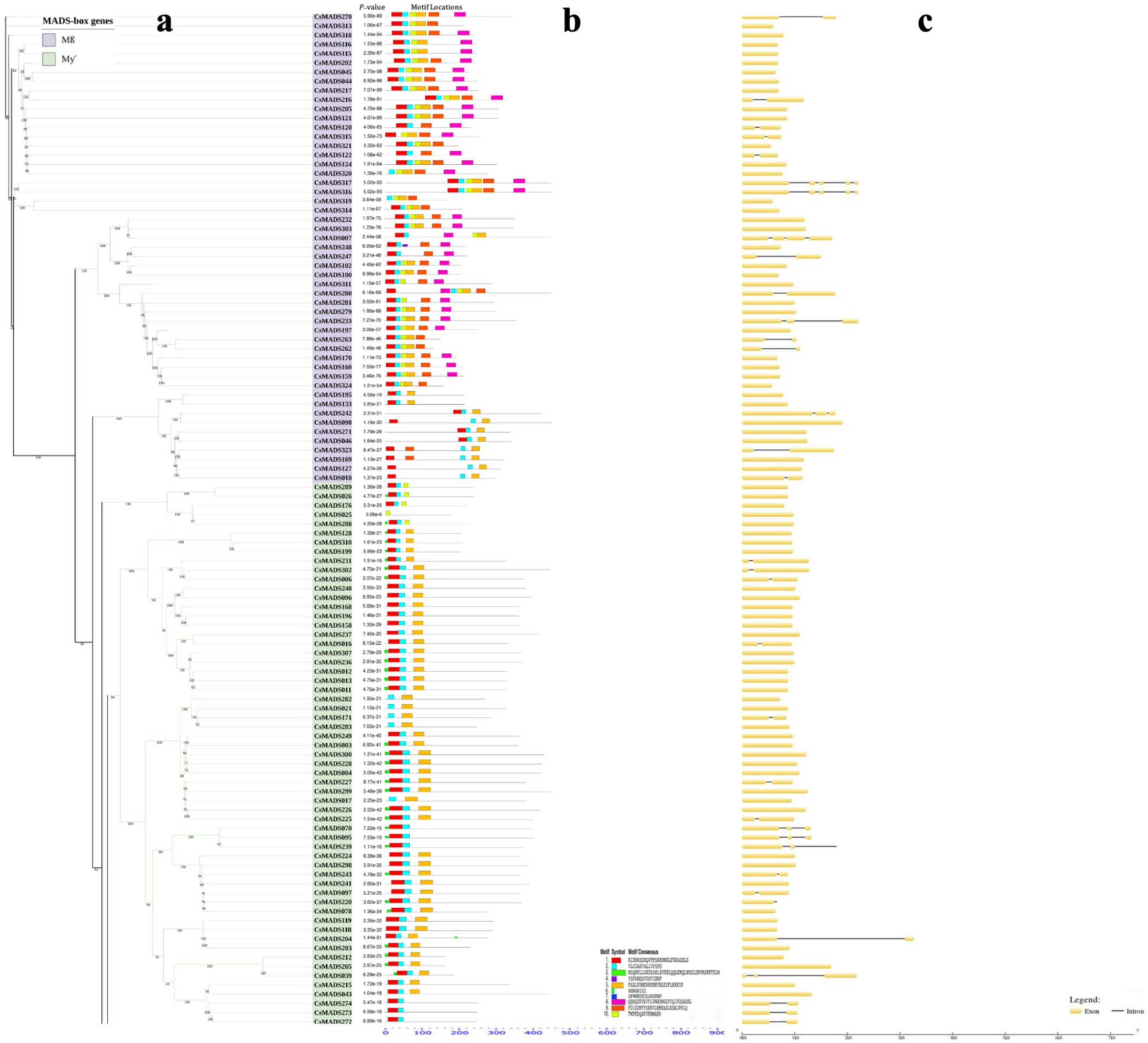


Fig. 3. Evolutionary tree, conserved motif, and gene structures of *Camelina sativa* MADS-box genes. (**a**) phylogenetic tree of Type I-MADS-box gene family (Ma and Mß) in *Camelina sativa*; protein sequences were used to construct the ML (maximum likelihood) tree. (**b**) Motif locations of the Type I-MADS-box (Ma and Mß) proteins. Different motifs are represented by different colored boxes. The box length represents motif length. (**c**) The exon-intron structure analyses of Type I-CsMADS-box gene family (Ma and Mß). The lengths of the exons and introns of each MADS-box genes are displayed proportionally. The yellow boxes represent exons, the black lines represent introns.

9 exhibit a predominance of Type II genes over Type I genes (Fig. 7b). Furthermore, *MADS-box* genes belonging to the same subclades frequently aggregate in particular regions of the chromosome. For instance, *CsMADS31* and *CsMADS32*, *CsMADS313* and *CsMADS34*, as well as *CsMADS318* and *CsMADS19*, which are associated with the FLC, Mβ, and TM3-like subclades respectively, are closely positioned on chromosome 2 (Fig. 7a).

Gene duplication analysis of *MADS-box* genes

To conduct an analysis of gene duplication, the coding sequences pertaining to the 325 *MADS-box* genes derived from *C. sativa* were employed as query sequences in BLAST searches against the complete set of *CsMADS-box* genes, utilizing an E-value threshold of $< 1e^{-10}$ and a sequence identity exceeding 85% (Table S3).

In evaluating the selection pressures exerted on duplicated *MADS-box* genes, we computed the ratio of non-synonymous (Ka) to synonymous (Ks) mutations (Ka/Ks) across 519 gene pairs. The results from the duplication analysis indicated that 230 out of the 325 *MADS-box* genes (70.77%) were present in two or more copies. Within this dataset, 235 *MADS-box* genes were found to have homologous counterparts located on duplicated genomic segments. Furthermore, a subset of 47 *MADS-box* genes was identified as exhibiting both segmental

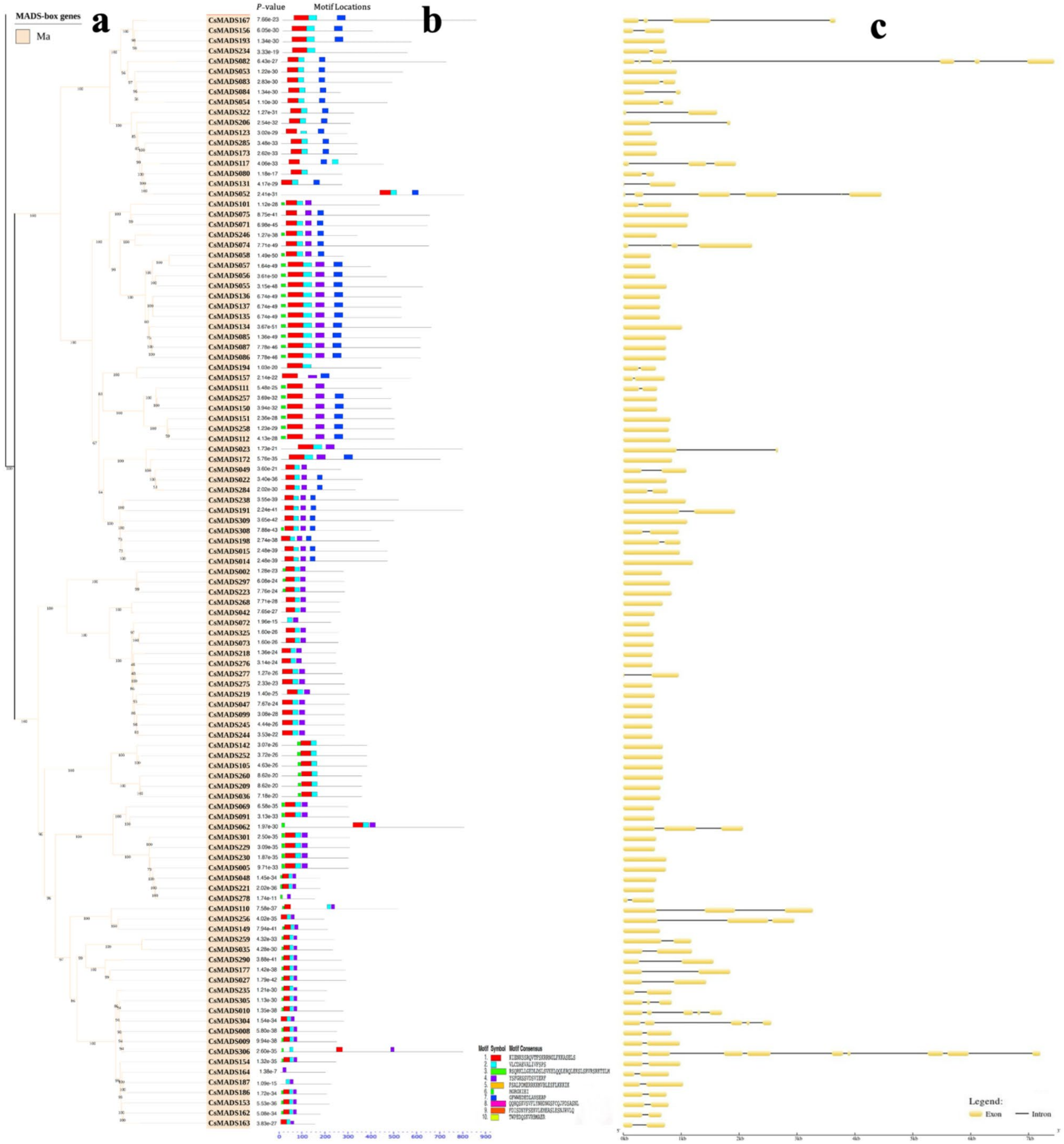


Fig. 4. Evolutionary tree, conserved motif, and gene structures of *Camelina sativa* MADS-box genes. (**a**) phylogenetic tree of Type I-MADS-box gene family (Mγ) in *Camelina sativa*; protein sequences were used to construct the ML (maximum likelihood) tree. (**b**) Motif locations of the Type I-MADS-box (Mγ) proteins. Different motifs are represented by different colored boxes. The box length represents motif length. (**c**) The exon-intron structure analyses of Type I-CsMADS-box gene family (Mγ). The lengths of the exons and introns of each MADS-box genes are displayed proportionally. The yellow boxes represent exons, the black lines represent introns.

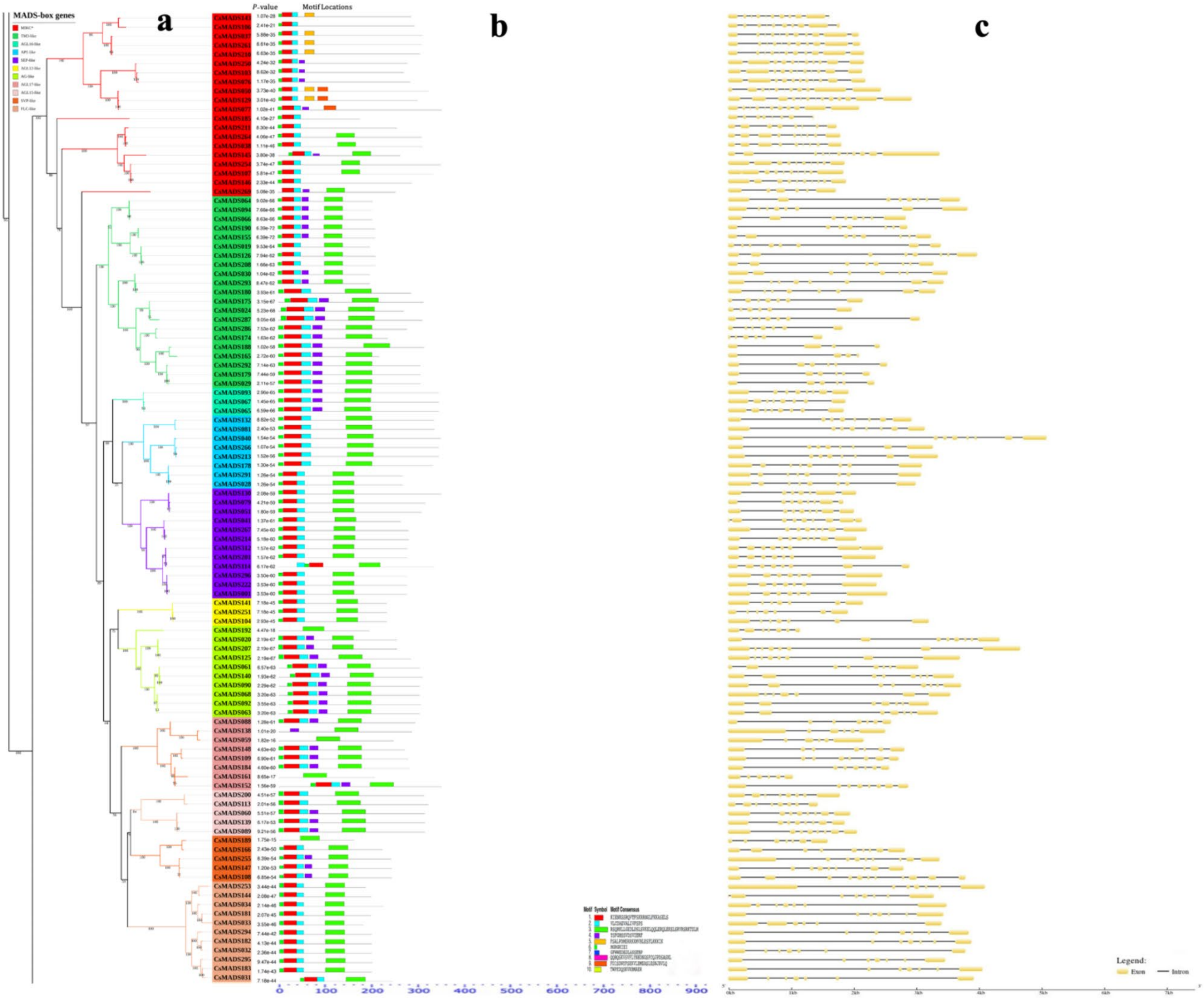


Fig. 5. Evolutionary tree, conserved motif, and gene structures of *Camelina sativa* MADS-box genes. (**a**) phylogenetic tree of Type II-MADS-box gene family in *Camelina sativa*; protein sequences were used to construct the ML (maximum likelihood) tree. (**b**) Motif locations of the Type I-MADS-box proteins. Different motifs are represented by different colored boxes. The box length represents motif length. (***c***) The exon-intron structure analyses of Type II-CsMADS-box gene family. The lengths of the exons and introns of each MADS-box genes are displayed proportionally. The yellow boxes represent exons, the black lines represent introns.

Motif	Width	Sequence
1	27	KIENKSSRQVTFSKRRNGLFKKASELS
2	15	VLCDAEVALIVFSPS
3	41	RSQRKLLGEDLDSLSVKELQQLERQLEKSLSRVRSRKTZLM
4	15	YSFGHSSVDSVIERF
5	25	FSALPDMERRKKMVBLESFLKKKIK
6	8	MGRGKIEI
7	15	GFWWEDEDLANSEBP
8	29	QQHQSKVSVFLYNHDNGSFCQJPDSASNL
9	27	FDISDNYFSEKVLEMEASLESNJRVLQ
10	15	TWPEDQSKVRBMAER

Table 2. Conserved motif of MADS-box gene family identified in *Camelina sativa* L.

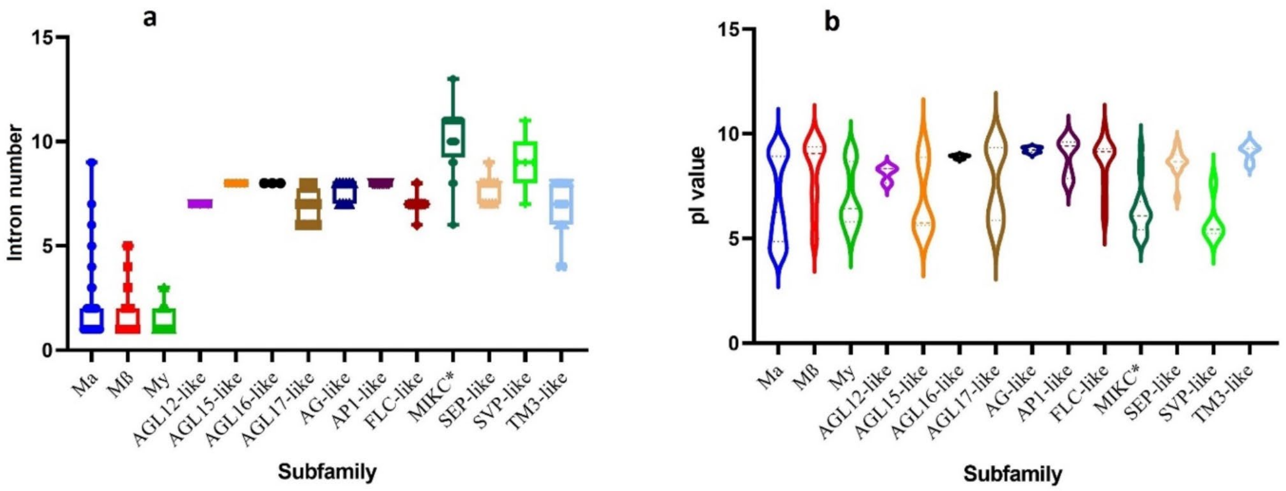


Fig. 6. Comparison between CsMADS-box subfamilies based on exon number (**a**) and pI (**b**).

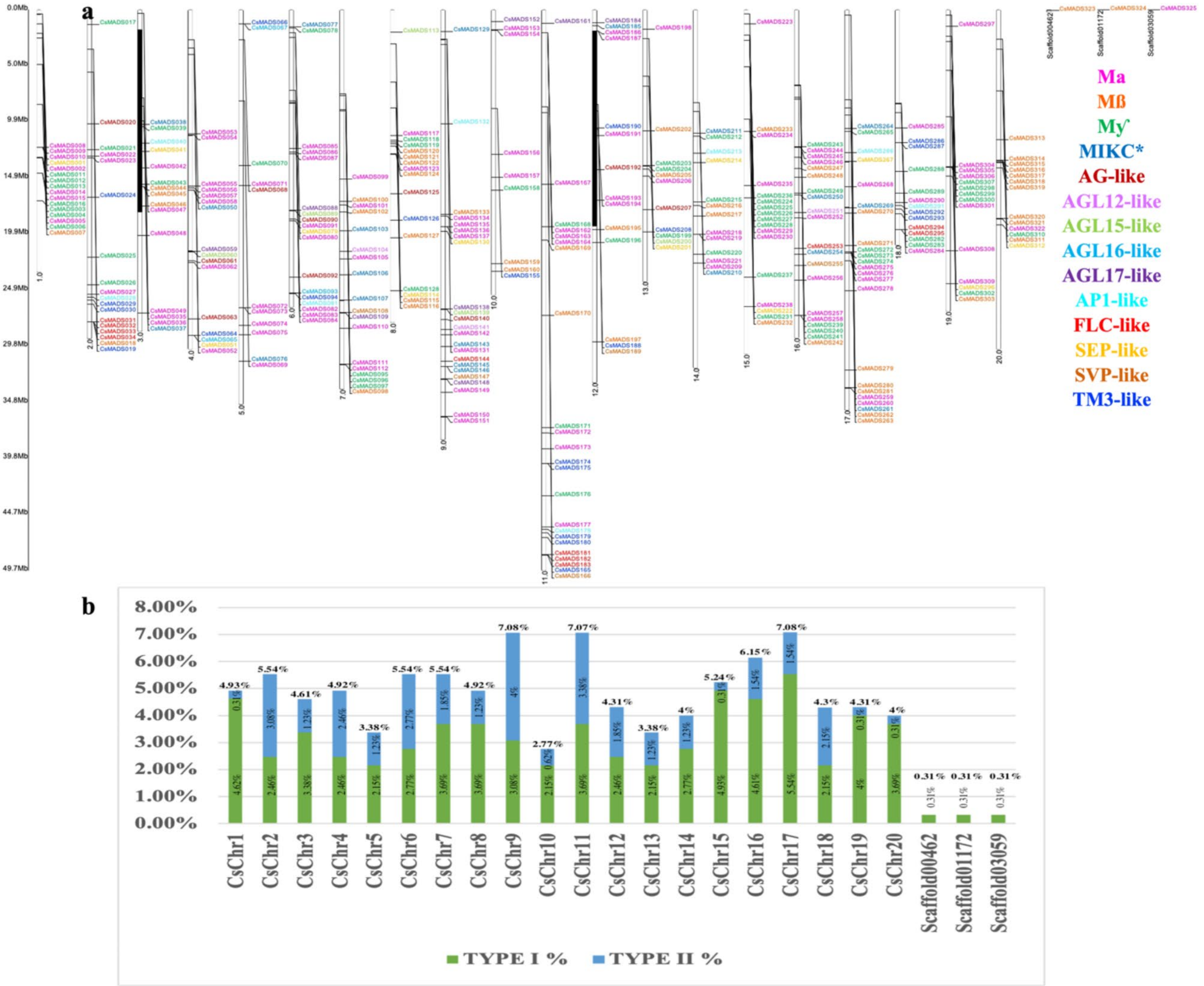


Fig. 7. Chromosomal location of *Camelina sativa* MADS-box genes. (**a**) The 325 CsMADS-box genes distributed on the 20 *Camelina sativa* chromosome. (**b**) The percentages of CsMADS-box genes on each chromosome.

and tandem duplications simultaneously. Conversely, the incidence of tandem duplications among the M-type *C. sativa MADS-box* genes was comparatively infrequent, with only 9 tandemly duplicated genes located on chromosomes 4, 17, and 20.

Promoter cis-regulatory element enrichment analysis of *MADS-box* genes

The promoter cis-regulatory element enrichment analysis performed on *MADS-box* genes yielded the identification of 39,111 putative cis-regulatory elements (CREs) within the cohort of 325 *CsMADS-box* genes. These CREs were systematically classified into six overarching categories based on their functional roles and responsiveness to various stimuli. Statistical evaluations demonstrated that promoter-associated elements constituted the most substantial fraction of all identified CREs (66%), followed by elements associated with developmental processes (13.1%), environmental stress-responsive elements (8.5%), hormone-responsive elements (5.6%), miscellaneous elements (5.5%), and site-binding-related elements (1.3%) (Table S4 and Fig. 8). The category pertaining to development-responsive elements encompassed 55 distinct types of cis-regulatory elements, representing the largest proportion of the overall dataset. This category included elements such as Box4, TCT-motif, GT1-motif, G-box, G-Box, GATA-motif (which is involved in light responsiveness), MYC (which is associated with cell-cycle regulation), AAGAA-motif (which is implicated in endosperm-specific negative expression), as1 (which is related to root-specific expression), CAT-box, and CCAAT-box (which is connected to meristem expression), O2-site (involved in the regulation of zein metabolism), circadian (which is relevant to circadian control), and RY-element (which is involved in seed-specific regulatory mechanisms). Eighteen distinct types of CREs were linked to plant hormone responsiveness, encompassing a diverse array of hormones, including ABRE (abscisic acid responsiveness), ERE (ethylene-responsive element), CGTCA-motif, CGTCA-motif, and TGACG-motif (MeJA-responsiveness), TCA-element and TCA (salicylic acid responsiveness), TGA-element and AuxRR-core (auxin-responsive element), GARE-motif, and p-box (gibberellins responsiveness). The third category comprised 17 types of CREs that are potentially responsive to environmental stresses, which included MYB, MYB-like sequences, Myb, Myb-binding sites, and MYB recognition sites (abiotic stress-responsive), ARE (which is crucial for anaerobic induction), WUN-motif, W-box, and WRE3 (involved in wounding and pathogen responses), TC-rich repeats (associated with defense and stress responsiveness), MBS (which is drought-inducible), LTR (which is responsive to low temperatures), DRE-core, and DRE1 (which are linked to dehydration, low-temperature, and salt stresses). The two categories of promoter-related elements and site-binding related elements encompassed eight and five types of CREs, respectively. The functional implications of the remaining 21 types of cis-regulatory elements remain to be elucidated (Table 3).

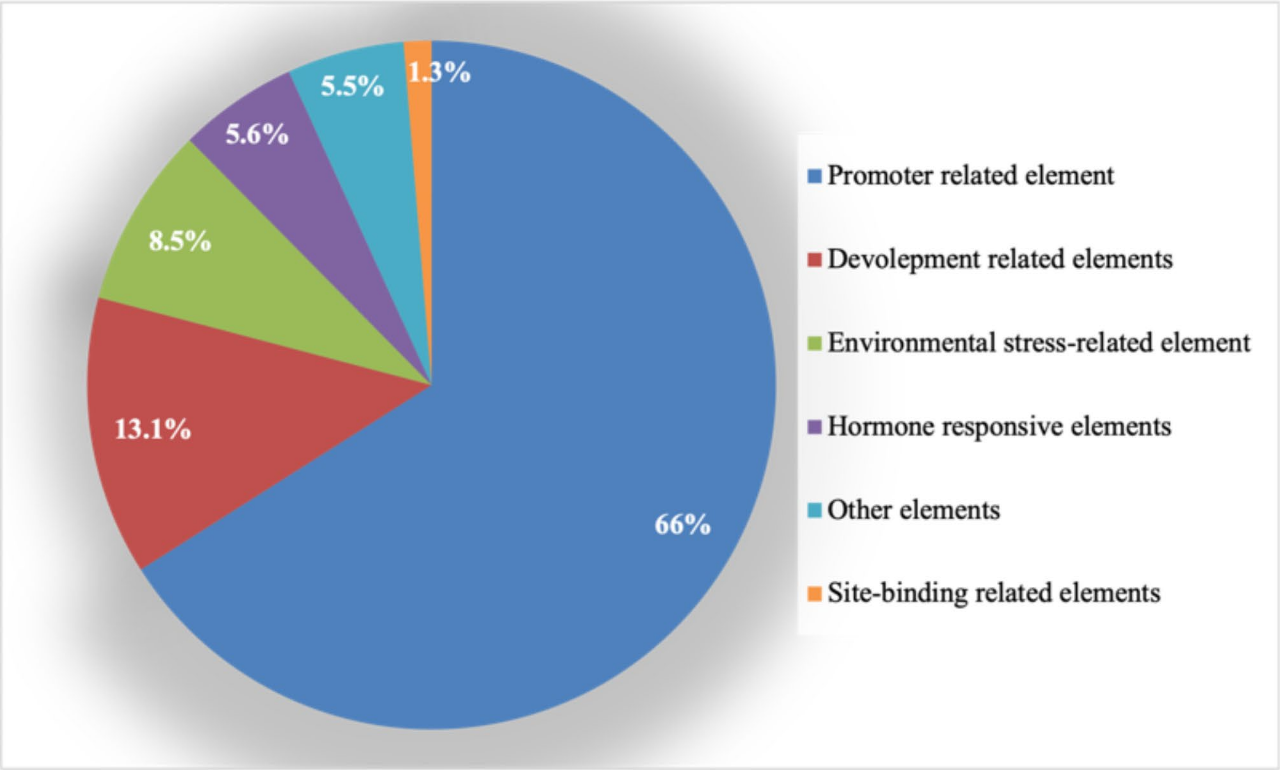


Fig. 8. Percentage distribution of cis-regulator elements (CREs) in the promoters of CsMADS-box genes based upon the putative functions.

Element name	Function	Total
TATA-box	core promoter element around – 30 of transcription start	14,796
CAAT-box	common cis-acting element in promoter and enhancer regions	8347
AT ~ TATA-box		2294
TATA		156
Unnamed_4	unknown	1756
Box4	part of a conserved module involved in light responsiveness	778
GT1-motif		423
TCT-motif		291
GATA-motif		169
AE-box		139
MYB	Abiotic stress responsive	771
MYB-like sequence		394
Myb		285
Myb-binding site		152
MYC	Regulates the expression of genes involved in cell-cycle control	750
Myc		159
ARE	Cis-acting regulatory element essential for the anaerobic induction	706
AAGAA-motif	involved in endosperm-specific negative expression	440
ABRE	cis-acting element involved in the abscisic acid responsiveness	431
ERE	ethylene-responsive element	373
CGTCA-motif	cis-acting regulatory element involved in the MeJA-responsiveness	319
TGACG-motif		301
as1	cis-acting regulatory element involved in the root-specific expression	314
G-box	cis-acting regulatory element involved in light responsiveness	280
G-Box		162
Unnamed_1	60 K protein binding site	251
STRE	Other elements	199
WUN-motif	wound-responsive element	198
W box	wounding and pathogen response	181
TC-rich repeats	cis-acting element involved in defense and stress responsiveness	137

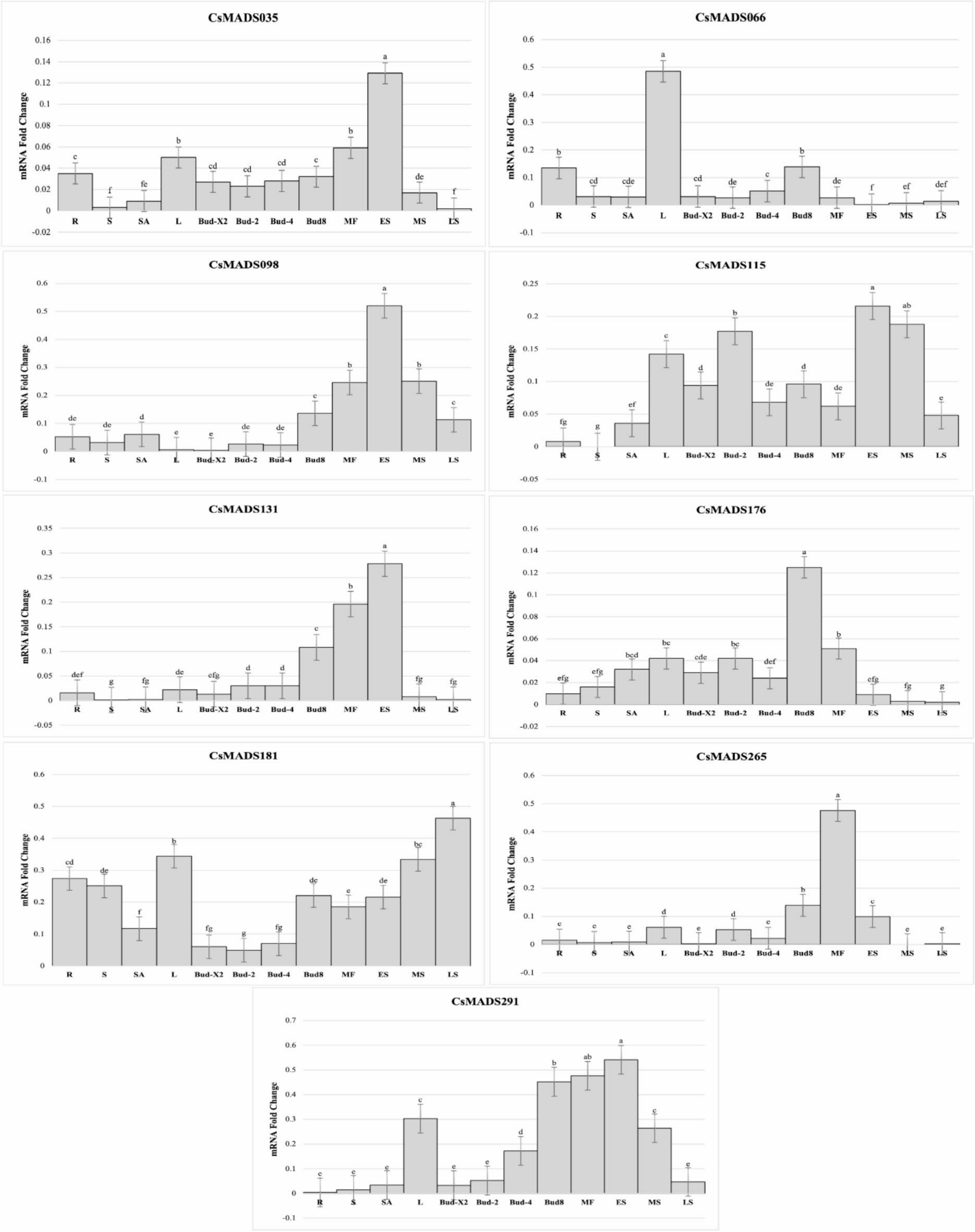
Table 3. Most frequent elements of the *MADS-box* genes in *Camelina sativa*.

Expression profiles of *CsMADS-box* genes in various organs and developmental stages

We conducted an investigation into the expression profiles of nine genes belonging to the *MADS-box* family in camelina, assessing their expression across various organs including roots, stems, leaves, shoots apex, flower buds (in four diameter sizes: < 2 mm (Bud-X2), ~ 2 mm (Bud-2), ~ 4 mm (Bud-4) and ~ 8 mm (Bud-8)), mature flowers and seeds (in three different stages: early seed development (ES), mid seed development (MS) and late seed development (LS)). The outcomes derived from the analysis of variance, which aimed to elucidate the expression patterns of the target genes in different organs, are depicted in the Fig. 9. The findings indicated a statistically significant disparity in gene expression across the various organs ($P < 0.05$). The results obtained from mean comparisons employing Duncan's test yielded the following insights:

CsMADS035 exhibited a markedly elevated expression level during the initial phase of seed development relative to other stages and organs (fold = 0.129), while it demonstrated the lowest expression during the late stage of seed development (fold = 0.002) and in roots (fold = 0.003). *CsMADS066* exhibited a significantly elevated expression level in leaves when compared to other organs (fold = 0.485) and demonstrated the lowest expression during the initial phase of seed development (fold = 0.002). *CsMADS098* revealed a markedly higher expression level during the initial phase of seed development compared to all other developmental stages and organs (fold = 0.520), with the lowest expression recorded in flower buds measuring less than 2 mm (fold = 0.004) and in leaves (fold = 0.006). *CsMADS115* showed a greater expression level in the early phase of seed development relative to other stages and organs (fold = 0.216) and displayed an absence of expression in the stem. *CsMADS131* demonstrated a higher expression level during the initial phase of seed development compared to other stages and organs (fold = 0.278), while exhibiting the lowest expression in the stem (fold = 0.001).

CsMADS176 presented a higher expression level during the 8 mm bud stage in comparison to other stages and organs (fold = 0.125), with the lowest expression occurring in the final phase of seed development (fold = 0.002). *CsMADS181* exhibited a higher expression level in the final phase of seed development compared to other organs (fold = 0.463), while showing the lowest expression during the 2 mm bud stage (fold = 0.049). *CsMADS265* demonstrated a higher expression level in mature flowers relative to other organs (fold = 0.476) and exhibited an absence of expression during the intermediate phase of seed development. *CsMADS291* showed


CsMADS035

CsMADS066

CsMADS098

CsMADS115

CsMADS131

CsMADS176

CsMADS181

CsMADS265

CsMADS291

Fig. 9. Real-time quantitative PCR expression analysis of *MADS-box* genes in various developmental stages of *C. sativa* (R: root, S: stem, SA: shoot apex, L: leaf, Bud-X2: < 2 mm, Bud-2: ~2 mm, Bud-4: ~4 mm, Bud-8: ~8 mm, MF: mature flower, ES: early seed development, MS: mid seed development and LS: Late seed development). The error bars represent the standard error of the means of three independent replicates. Values denoted by the same letter did not differ significantly at $P < 0.05$ according to Duncan's multiple range tests.

a significantly higher expression level during the early stages of seed development (fold = 0.541) and in mature flowers (fold = 0.476) compared to other organs, while presenting the lowest expression in the root (fold = 0.004).

Expression patterns of *CsMADS-box* genes in response to drought stress

We further investigated the expression of selected genes subjected to drought stress conditions. The results stemming from the analysis of variance aimed at assessing gene expression under drought stress are presented in the Fig. 10. The findings revealed a significant difference in gene expression across varying stress conditions ($P < 0.05$). The outcomes from the mean comparisons utilizing Duncan's test revealed the following results:

The expression levels of *CsMADS035* in leaves at 4 h post-stress exposure in DS3 and DS2 conditions were significantly different from the control, with increases of 11.6-fold and 7.31-fold, respectively (the disparity in expression between the two treatments was significant). No significant differences in gene expression were recorded during other time points of stress application. The expression levels of *CsMADS066* in leaves, measured four hours post-stress, exhibited a reduction of 5.05-fold and 4.8-fold in DS3 and DS2, respectively, when contrasted with the control (the differential expression between the two experimental conditions was not statistically significant). At the 8-hour mark, the gene expression in DS2 and DS3 demonstrated an enhancement of 2.93-fold and 2.43-fold relative to the control, respectively (the differential expression between the two experimental conditions was not statistically significant). Subsequently, at the 24-hour stress, the expression of the gene in DS3 and DS2 increased by 7.13-fold and 3-fold in relation to the control, respectively (the differential expression between the two experimental conditions was statistically significant). At the final hour of stress, the expression of the gene in DS2 and DS3 declined by 27.05-fold and 5.94-fold compared to the control, respectively (the differential expression between the two experimental conditions was not statistically significant).

The expression levels of *CsMADS098* in leaves, 8 h subsequent to stress in DS3 and DS2, decreased by 4.35-fold and 1.54-fold, respectively, in comparison to the control (the differential expression between the two experimental conditions was statistically significant). The most pronounced reduction in expression (20-fold) was observed in relation to DS3 stress at the 24-hour mark. The expression of *CsMADS115* in leaves, 4 h following stress in DS3 and DS2, increased by 3.11-fold and 1.69-fold compared to the control, respectively. At the 8-, 12-, and 24-hour intervals, a trend of diminishing gene expression was discerned in DS3 when contrasted with the control (8.5-fold, 2.6-fold, and 7.2-fold, respectively). At 48-hours of stress in DS3, a resurgence in gene expression was documented (2.44-fold). In the context of DS2 stress, the pattern of gene expression was irregular, as evidenced by a decrease at eight hours (1.61-fold), an increase at 12 h (1.53-fold), a decrease at 24 h (3.74-fold), and a subsequent increase at forty-eight hours (1.39-fold) in gene expression.

The expression levels of *CsMADS131* in leaves, 4 h post-stress in DS3 and DS2, exhibited a decline of 5.31-fold and 1.81-fold relative to the control, respectively. Throughout the 8-, 12-, 24-, and 48-hour intervals, a consistent trend of decreasing gene expression was observed in both DS2 and DS3 when compared to the control. The most substantial reduction in expression (7.2-fold) was associated with DS2 stress at the 24-hour mark. The expression levels of *CsMADS176* in leaves, exhibited a marked increase up to 8 h post-stress in both DS3 and DS2 when contrasted with the control group, with the peak expression observed at 8 h yielding 7.03-fold and 5.3-fold increases, respectively. At the 12- and 48-hour time points, no statistically significant differences were detected, while at the 24-hour mark, a significant reduction in gene expression was noted exclusively in DS3 (10-fold).

The expression of *CsMADS181* in leaves, at 4-, 8-, 12-, and 24 h following stress application in DS3 exhibited a significantly decreasing trend in comparison to the control, with the most pronounced reduction in expression (14.6-fold) occurring at the 24-hour time point. A notable decreasing trend in gene expression for DS2 was recorded at 8, 12, 24, and 48 h, with minimal gene expression (7.3-fold) observed at 12 h. The expression of *CsMADS265* in leaves, at 4-, 8-, 12-, and 24 h subsequent to stress in DS3 demonstrated a significant increase relative to the control, with the most substantial increase in expression (34.67-fold) occurring at the 12-hour interval. Furthermore, a significant elevation in gene expression within DS2 was observed at 12- and 48 h, with the peak gene expression (15.4-fold) recorded at the 12-hour mark. The expression of *CsMADS291* in leaves, at 4- and 48 h following stress in DS3, displayed a significant decreasing trend when compared to the control, with the lowest gene expression (5-fold) noted at the 48-hour time point. Conversely, within the same stress context, a significant increase in gene expression was documented at 8-, 12-, and 24 h, with the highest level of gene expression (4.36-fold) occurring at 8 h. Moreover, a significant increase in gene expression within DS2 was also observed at 8-, 12-, and 24 h, with the peak gene expression (2.5-fold) recorded at the 8-hour interval.

Discussion

Comprehensive identification of an extensive family of *MADS-box* genes in camelina

In the present investigation, bioinformatic analysis identified a total of 325 *MADS-box* genes categorized into Type I and Type II within the genome of *Camelina sativa*. When contrasted with prior research, the quantity of *MADS-box* genes demonstrates significant variability between monocotyledonous and dicotyledonous species. For example, *Oryza sativa* L. (75), *Zea mays* L. (75), and *Sorghum bicolor* (65)[45,58] exhibit a reduced number of *MADS-box* genes in comparison to Arabidopsis (111), *Solanum tuberosum* (156), and *Solanum lycopersicum* (131)[59,60]. This finding indicates a probable reduction in the number of *MADS-box* genes within monocotyledonous plants over the course of evolutionary history. In order to investigate the evolutionary dynamics of *MADS-box* genes in *C. sativa*, a phylogenetic tree was constructed utilizing the amino acid sequences of MADS-box proteins derived from both *C. sativa* and *A. thaliana* (Figs. 1 and 2). The analysis revealed that *C. sativa* possesses a substantially higher number of Type I *MADS-box* genes (219) compared to *foxtail millet* (43), *Setaria viridis* (43), and rice (45)[10,45]. On the other hand, the number of Type II *MADS-box* genes was relatively consistent across these species. This pattern suggests that monocotyledons may have predominantly lost Type I *MADS-box* genes during their evolutionary history.

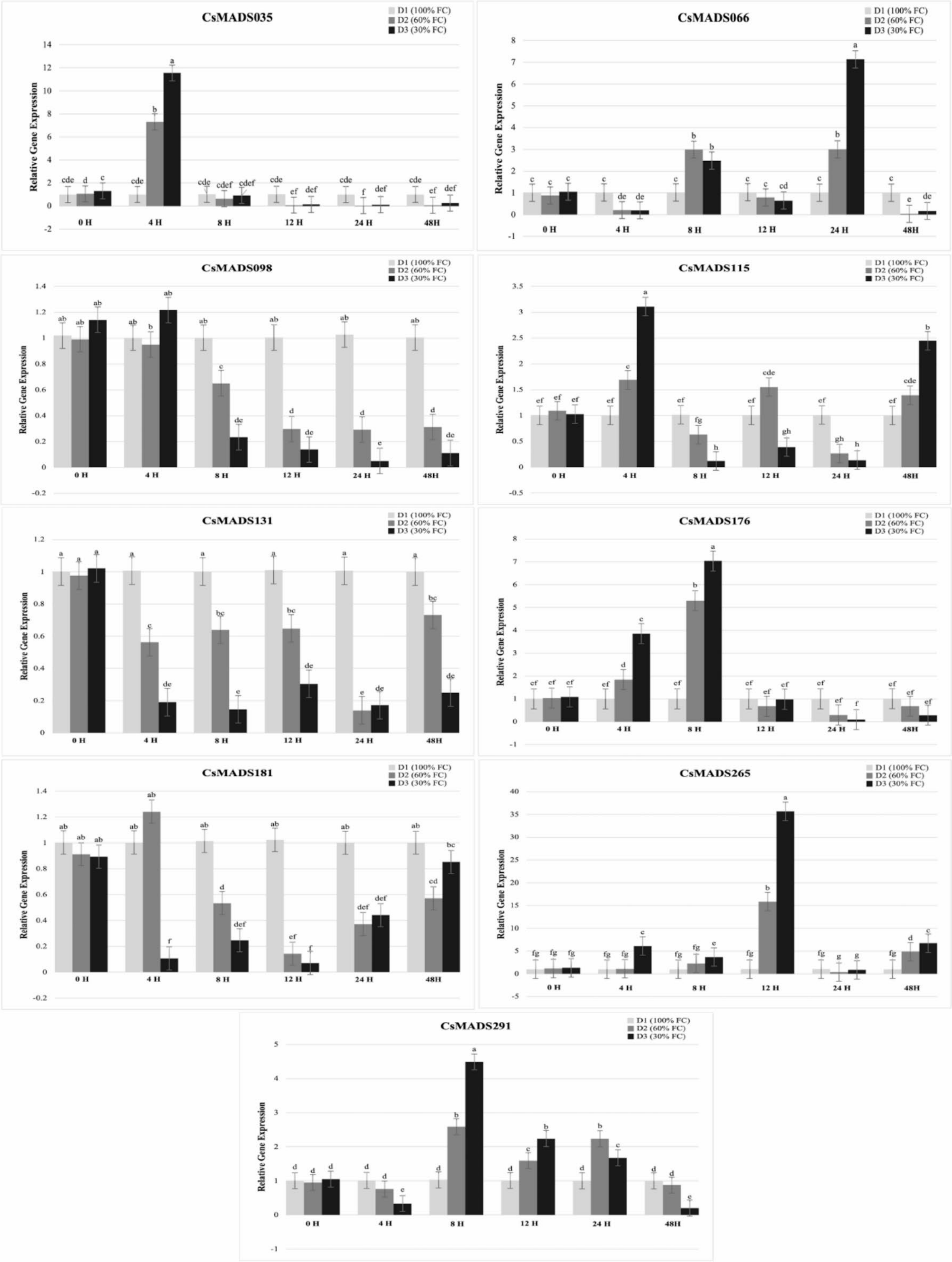


Fig. 10. Real-time quantitative PCR expression analysis of *MADS-box* genes after drought stress treatment (0–48 h) in *C. sativa* (DS1: 100% field capacity, DS2: 60% field capacity and DS3: 30% field capacity). The error bars represent the standard error of the means of three independent replicates. Values denoted by the same letter did not differ significantly at $P < 0.05$ according to Duncan's multiple range tests.

Within *C. sativa*, the Type II *MADS-box* genes are classified into the MIKC^C and MIKC* clades, with the MIKC^C clade further subdivided into ten distinct subclades. In general, MADS-box proteins exhibiting analogous motifs tend to congregate in phylogenetic analyses (the AP1, SEP-like, and AGL12-like subclades), implying that constituents of the same subclade likely perform analogous functions[50]. Nevertheless, these distinctions may be attributable to the evolutionary pathway of MADS-box proteins in *C. sativa* (Figs. 3b and 4b, and 5b).

The structural variability of genes is of paramount importance in the evolutionary processes of multigene families. Investigations reveal that during the evolutionary history of plant *MADS-box* genes, occurrences of intron loss and insertion mutations are prevalent phenomena[7,59]. Our analysis specifically observed a notable discrepancy in the intron count among *CsMADS-box* genes, ranging from 0 to 12. In *C. sativa*, Type I *MADS-box* genes predominantly lack introns, with a few exceptions, whereas Type II genes generally encompass at least three introns, the highest number being identified within the MIKC^C subgroup (Figs. 3c and 4c, and 5c). Furthermore, an examination of protein motifs within the same gene family displayed non-homogeneous patterns. These variations in intron presence among *MADS-box* genes within the same subgroup imply that intron gain or loss may represent a recurrent evolutionary phenomenon in *CsMADS-box* genes, potentially contributing substantially to the functional diversity observed within the *CsMADS-box* family. Despite these discrepancies, the majority of *CsMADS-box* genes within the same group demonstrated analogous intron-exon arrangements, similar to those found in Type I and Type II members of the *CsMADS-box* family, and their predicted proteins exhibited comparable sets of motifs. Comparable evolutionary patterns have been documented in monocot species such as rice[45] and wheat[61], underscoring the evolutionary conservation of the *MADS-box* family across diverse plant taxa.

Gene duplication appears to have been a significant driving force behind the variation of the *MADS-box* genes family

The occurrence of gene duplication has played a crucial role in the expansion of *MADS-box* genes within the plant genus[24]. In *C. sativa*, our analysis indicates that the significant increase in *MADS-box* genes arose from both tandem and segmental duplications but mainly due to segmental duplication (Table S3), and higher frequencies of segmental duplications led to the generation of many *MADS-box* gene homologs across all chromosomes of *C. sativa*.

No substantial gene clusters or regions of high density for *CsMADS-box* genes were discovered, potentially attributable to the infrequent incidence of tandem duplications[62]. Evolutionary analysis of *C. sativa* supported these findings, with only 2.77% of *C. sativa* genes being tandem duplicates, in stark contrast to 27% of *Arabidopsis* genes within a 100-kbp genomic interval[63].

Despite our findings, the expansion of the plant *MADS-box* genes family is primarily attributed to tandem duplication events[61,64]. As mentioned aboved, the analysis revealed that *C. sativa* possesses a substantially higher number of Type I *MADS-box* genes (219) compared to *foxtail millet* (43), *Setaria viridis* (43), and rice (45); suggesting that camelina genome undewent more gene duplication than the *foxtail millet*, *Setaria viridis*, and rice[10,45].

We determined the divergence times for closely related CsMADS-box gene pairs based on the phylogenetic tree (mutation rate of 8.22×10^{-9} substitutions/synonymous site/year for Brassicaceae species) (Kagale et al., 2014). The majority of gene pairs (98%) exhibited divergence roughly 28 million years ago (MYA), accompanied by a standard deviation of 6 MYA, which suggests that these duplication events occurred prior to the divergence of Brassicaceae species[65]. Furthermore, the Ka/Ks ratios (Ka/Ks ratio <1) imply that these genes have predominantly experienced purifying selection throughout their evolutionary trajectory[53]. This implies that the dynamic changes in the *CsMADS-box* genes family have likely played a role in the plant's adaptation to various environmental conditions, aiding in its widespread ecological distribution[66]. Post-duplication, genes often undergo sub-functionalization, where they split existing roles, or neo-functionalization, where they acquire new roles, thereby enhancing plant adaptability[67,68].

Cis-regulatory elements showed their key role in plant growth, development, and resilience

Cis-regulatory elements (CREs) are non-coding DNA sequences predominantly situated upstream of genes. They are recognized by transcription factors that modulate gene expression in response to a variety of environmental conditions[69]. CREs are integral to regulatory networks as they enable interactions between DNA and regulatory proteins. Through the examination of the promoter region (1.5 kbp upstream of the start codon), critical *cis*-elements within *CsMADS-box* genes were identified (Table S4 and Fig. 8). The investigation of these *cis*-elements within promoters indicates that *CsMADS-box* genes are implicated in the regulation of plant growth and development in reaction to diverse environmental stresses and stimuli[6,45,70]. This understanding lays the groundwork for subsequent functional characterization of *MADS-box* gene family constituents in *C. sativa* and other agriculturally significant crops.

Organ developmental regulation of *MADS-box* genes

Investigations into the genetic and molecular frameworks governing floral development in the model eudicot, *Arabidopsis thaliana*, have underscored the essential function of *MADS-box* genes in determining floral organ identity[29,71]. Notably, the expression of various *MADS-box* genes beyond the floral context suggests that this gene family also plays a role in vegetative development, encompassing functions in root growth and leaf development[15,24]. Additionally, the existence of *MADS-box* genes in non-flowering plant lineages such as gymnosperms and mosses underscores that the roles of these genes extend far beyond the development of reproductive organs to encompass broader regulatory functions in plant growth and differentiation[32,72].

In camelina, an *AP1-like homologous* gene (*CsMADS291*) was identified, exhibiting high expression during the early stages of seed development, bud8 stage, and in leaves (Fig. 10). AP1-like functions as a floral organ identity gene, promoting the development of petals and sepals while defining floral meristem identity[71,73]. These findings indicate that this gene, in addition to its involvement in flower primordium development, also contributes to leaf development and the early stages of seed development.

The expression of the *FLC-like homologous* gene (*CsMADS181*) was observed throughout the majority of vegetative phases, as well as during inflorescence and seed development (Fig. 9). It appears that the transcripts of *CsMADS181* exhibited downregulation during the early stages of bud development, akin to their counterparts in *Arabidopsis*[74]. In both *Arabidopsis* and the Brassicaceae, FLC functions as a flowering inhibitor and serves as a critical integration point for both environmental and endogenous pathways that modulate flowering time through vernalization and autonomous mechanisms in *Arabidopsis*[75]. These observations imply that a decline in the expression of *CsMADS181* during the preliminary phases of flower bud development may act as a significant trigger for the camelina's transition into its reproductive phase. Subsequently, the elevation of its expression during the mid-seed development phase indicates its role in seed growth.

In camelina, the *TM3-like homologous* gene (*CsMADS066*) was identified and demonstrated high expression levels in the vegetative organs (root and leaf) (Fig. 9). *TM3-like* genes within *Arabidopsis* have been documented to play roles in root development[24,76]. Furthermore, these genes are preferentially expressed in the vegetative structures of various plant species[77,78]. Such findings suggest that this gene is integral in facilitating the vegetative stage of camelina, particularly in the development of leaves and roots.

Additionally, we investigated the expression profiles of six genes from the Type I *MADS-box* family in camelina. Both Mα-type (*CsMADS035* and *CsMADS131*) and Mß-type (*CsMADS098* and *CsMADS115*) genes exhibited peak expression during the nascent stages of seed development (Fig. 9). In both *Arabidopsis* and rice, Mα-type and Mß-type *MADS-box* genes demonstrate analogous expression patterns. Consequently, our findings may signify the involvement of these genes in seed development[79–81]. However, a notable observation was that Mγ-type (*CsMADS176* and *CsMADS265*) genes showed maximal expression during the 8 mm bud and mature flower stages, which deviates from the behavior exhibited by the Type I *MADS-box* genes and aligns more closely with the expression profile of Type II *MADS-box* genes (Fig. 9). Given that Type I (Mγ-type) *MADS-box* genes have been previously associated with seed development, our results may not correspond with earlier research findings[16,82].

Drought stress differentially affected *MADS-box* gene expression

Drought is a key factor that hinders plant growth and survival, leading plants to develop diverse strategies to cope with limited water availability[83,84]. Drought induces response mechanisms that involve both ABA-dependent and ABA-independent signal transduction pathways[85]. To understand the response of MADS-box family genes to drought stress, we examined the expression profile of 9 genes under drought stress conditions. The overall trend of gene expression in *CsMADS035*, *CsMADS066*, *CsMADS115*, *CsMADS176*, *CsMADS265*, and *CsMADS291* was an increase in expression, which showed more up-upregulation in DS3 treatment (Fig. 10). In *CsMADS098*, *CsMADS131*, and *CsMADS181* the expression trend was a decrease, which showed more down-upregulation in DS3 treartment (Fig. 10). In alignment with the latter, in *Rhododendron hainanense*, *RhMADS24*, *RhMADS25*, *RhMADS39*, and *RhMADS44* were significantly modulated under conditions of temperature and waterlogging stress[86]. In *Hevea brasiliensis*, 12 *HbMADS-box* genes were identified as being associated with floral development, with all of these floral-enriched *HbMADS-box* genes being regulated by hormonal influences, as well as by salt, cold, high-temperature, and drought stressors[87]. In rice, the expression levels of four *OsMADS-box* genes were markedly upregulated in response to low temperature and dehydration stress. Additionally, three *OsMADS-box* genes were found to play negative regulatory roles under conditions of dehydration and salt stress[45]. Furthermore, Lee and colleagues elucidated that *OsMADS26* exerts a positive regulatory influence on numerous genes associated with stress-related mechanisms[13]. Our findings imply that *MADS-box* genes could be integral to the abiotic stress responses in plants via cis-regulatory elements (CREs) (Fig. 8) and various factors that engage distinct signaling pathways. To date, none of these genes have undergone functional characterization utilizing diverse methodologies; thus, an investigation into the functionality of this gene family, alongside the elucidation of the role of *CsMADS-box* genes in conferring stress resistance, could yield significant theoretical insights and foundational knowledge.

Conclusion

Camelina serves as an ideal crop for cultivation in suboptimal soils and arid regions where other agricultural species exhibit challenges in growth. The application of an array of bioinformatics tools in plant and biological research significantly augments our comprehension of biological systems. This investigation constitutes the inaugural comprehensive analysis of the MADS-box transcription factor family in *C. sativa* L. We provide a meticulous account of the *MADS-box* genes identified in *C. sativa*, encompassing their biochemical properties, phylogenetic relationships, chromosomal positioning, gene duplication assessments, conserved protein motifs, gene architectures, promoter elements, and transcript-level gene expression analyses. Transcriptomic data coupled with quantitative real-time fluorescence outcomes suggest that certain *CsMADS-box* genes are pivotal in mediating drought stress responses. The results underscore the critical involvement of *MADS-box* genes in the reaction to environmental disturbances. The identification of candidate genes paves the path for subsequent functional characterizations and the potential application of these genes in enhancing drought tolerance through contemporary biotechnological methodologies.

Data availability

The data generated or analyzed in this study are included in this article. Other materials that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 6 January 2025; Accepted: 10 March 2025

References

1. Castelán-Muñoz, N. et al. MADS-box genes are key components of genetic regulatory networks involved in abiotic stress and plastic developmental responses in plants. *Front. Plant Sci.* **10**, Preprintathttpsdoiorg103389fpls201900853 (2019).
2. Sarvepalli, K., Das Gupta, M., Challa, K. R. & Nath, U. Molecular cartography of leaf development — role of transcription factors. *Curr. Opin. Plant. Biol.* **47**, 22–31 (2019).
3. Thangavel, G. & Nayar, S. A survey of MIKC type MADS-box genes in non-seed plants: algae, bryophytes, lycophytes and ferns. *Front. Plant Sci.* **9**, Preprintathttpsdoiorg103389fpls201800510 (2018).
4. Wang, L. N., Fan, S. L. & Song, M. Z. Advances in the research of MADS-box gene in plant. *Biotechnol. Bull.* **08**, 12–19 (2010).
5. He, C., Si, C., Teixeira Da Silva, J. A., Li, M. & Duan, J. Genome-wide identification and classification of MIKC-type MADS-box genes in streptophyte lineages and expression analyses to reveal their role in seed germination of Orchid. *BMC Plant. Biol.* **19** 1–15, (2019).
6. Saha, G. et al. Genome-wide identification and characterization of MADS-box family genes related to organ development and stress resistance in brassica Rapa. *BMC Genom.* **16**, 1–21 (2015).
7. Sheng, X. G. et al. Genome wide analysis of MADS-box gene family in brassica Oleracea reveals conservation and variation in flower development. *BMC Plant. Biol.* **19**, (2019).
8. Wei, B. et al. Genome-wide analysis of the MADS-Box gene family in brachypodium distachyon. *PLoS One* **9**, (2014).
9. Won, S. Y., Jung, J. A. & Kim, J. S. Genome-wide analysis of the MADS-Box gene family in chrysanthemum. *Comput. Biol. Chem.* **90**, 107424 (2021).
10. Zhao, W. et al. Genome-Wide analysis of MADS-Box genes in Foxtail millet (Setaria Italica L.) and functional assessment of the role of SiMADS51 in the drought stress response. *Front. Plant. Sci.* **12**, (2021).
11. Zhang, Z. et al. Characterization and expression analysis of six MADS-box genes in maize (Zea Mays L). *J. Plant. Physiol.* **169**, 797–806 (2012).
12. Guo, X. et al. Solanum lycopersicum agamous-like MADS-box protein AGL15-like gene, SlMBP11, confers salt stress tolerance. *Mol. Breed.* **36**, 1–15 (2016).
13. Lee, S. et al. Further characterization of a rice AGL12 group MADS-Box gene, OsMADS26. *Plant. Physiol.* **147**, 156–168 (2008).
14. Boden, S. A. & Østergaard, L. How can developmental biology help feed a growing population? *Dev. (Cambridge)* 146, (2019).
15. Theißen, G. Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant. Biol.* **4**, 75–85 (2001).
16. Köhler, C. et al. The polycomb -group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev.* **17**, 1540–1553 (2003).
17. Day, R. C., Herridge, R. P., Ambrose, B. & Macknight, R. C. Transcriptome analysis of proliferating Arabidopsis endosperm reveals biological implications for the control of syncytial division, cytokinin signaling, and gene expression regulation. *Plant. Physiol.* **148**, 1964–1984 (2008).
18. Wuest, S. E. et al. Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr. Biol.* **20**, 506–512 (2010).
19. Passmore, S., Maine, G. T., Elble, R., Christ, C. & Tye, B. K. Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MATα cells. *J. Mol. Biol.* **204**, 593–606 (1988).
20. Sommer, H. et al. Deficiens, a homeotic gene involved in the control of flower morphogenesis in antirrhinum Majus: the protein shows homology to transcription factors. *EMBO J.* **9**, 605–613 (1990).
21. Yanofsky, M. F. et al. The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature* **346**, 35–39 (1990).
22. Norman, C., Runswick, M., Pollock, R. & Treisman, R. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* **55**, 989–1003 (1988).
23. Honma, T., Goto, K., Carruthers, K., Johnson, K. & Urban, J. Complexes of MADS-Box proteins are sufficient to convert leaves into floral organs. (2010).
24. Alvarez-Buylla, E. R. et al. MADS-box gene evolution beyond flowers: Expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J.* **24**, 457–466 (2000).
25. Gramzow, L. & Theissen, G. A Hitchhiker's guide to the MADS world of plants. *Genome Biol.* **11**, 214–225 (2010).
26. Kaufmann, K., Melzer, R. & Theißen, G. MIKC-type MADS-domain proteins: Structural modularity, protein interactions and network evolution in land plants. *Gene* **347**, 183–198 (2005).
27. De Bodt, S., Raes, J. & Van De Peer, Y. And then there were many: MADS goes genomic. *Trends Plant. Sci.* **8**, 475–483 (2003).
28. Mcgonigle, B., Bouhidel, K. & Irish, V. F. Nuclear localization of the Arabidopsis APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression. *Genes Dev.* **10**, 1812–1821 (1996).
29. Riechmann, J. DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Res.* **24**, 3134–3141 (1996).
30. Yang, Y. & Jack, T. Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. *Plant. Mol. Biol.* **55**, 45–59 (2004).
31. Puranik, S. et al. Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in Arabidopsis. *Plant. Cell.* **26**, 3603–3615 (2014).
32. Henschel, K. et al. Two ancient classes of MIKC-type MADS-box genes are present in the moss physcomitrella patens. *Mol. Biol. Evol.* **19**, 801–814 (2002).
33. Dong, J. J. et al. Gene expression and phylogenetic analysis of MADS-box family genes in Ginkgo Biloba. *Zhiwu Shengli Xuebao/Plant Physiol. J.* **54**, 1055–1063 (2018).
34. Xiang, L., Chen, Y., Chen, L. P. & Sun, C. B. Expression analysis of B, C and E class MADS-box genes in cymbidium ensifolium. *Acta Horticulturae Sinica*. **45**, 1595–1604 (2018).
35. Kong, W. L., Zhang, K. D. & Wu, J. C. Genome-wide identification and phylogenetic analysis of MADS-box family gene in beta vulgaris. *Acta Agriculturae Boreali-Sinica*. **33**, 86–95 (2018).
36. Kagale, S. et al. The emerging biofuel crop camelina sativa retains a highly undifferentiated hexaploid genome structure. *Nat. Commun.* **5**, (2014).
37. Vollmann, J. & Eynck, C. Camelina as a sustainable oilseed crop: Contributions of plant breeding and genetic engineering. Biotechnology Journal vol. 10 525–535 Preprint at (2015). https://doi.org/10.1002/biot.201400200
38. Ghamkhar, K. et al. Camelina (*Camelina sativa* (L.) Crantz) as an alternative oilseed: molecular and ecogeographic analyses. *Genome* **53**, 558–567 (2010).
39. Amyot, L. et al. Assessment of antinutritional compounds and chemotaxonomic relationships between camelina sativa and its wild relatives. *J. Agric. Food Chem.* **67**, 796–806 (2019).
40. Zanetti, F. et al. The opportunities and potential of camelina in marginal land in Europe. *Ind. Crops Prod.* **211**, 118224 (2024).
41. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145 (2018).

42. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
43. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
44. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
45. Arora, R. et al. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genom.* **8**, 242 (2007).
46. Hu, B. et al. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**, 1296–1297 (2015).
47. Chen, C. et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant.* **13**, 1194–1202 (2020).
48. Liu, C. et al. Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (Ananas comosus). *BMC Genom.* **18**, 503 (2017).
49. Gu, Z., Cavalcanti, A., Chen, F. C., Bouman, P. & Li, W. H. Extent of gene duplication in the genomes of drosophila, nematode, and yeast. *Mol. Biol. Evol.* **19**, 256–262 (2002).
50. Duan, W. et al. Genome-wide analysis of the MADS-box gene family in brassica Rapa (Chinese cabbage). *Mol. Genet. Genomics.* **290**, 239–255 (2015).
51. Tang, J., Wang, F., Hou, X. L., Wang, Z. & Huang, Z. N. Genome-Wide fractionation and identification of WRKY transcription factors in Chinese cabbage (Brassica Rapa Ssp. pekinensis) reveals collinearity and their expression patterns under abiotic and biotic stresses. *Plant. Mol. Biol. Rep.* **32**, 781–795 (2014).
52. Hurst, L. D. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* **18**, 486–487 (2002).
53. Zhang, Z. et al. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteom. Bioinf.* **4**, 259–263 (2006).
54. Zhang, L. et al. Genome-wide analysis and expression profiling under heat and drought treatments of HSP70 gene family in soybean (Glycine max L). *Front. Plant. Sci.* **6**, (2015).
55. Chao, W. S., Wang, H., Horvath, D. P. & Anderson, J. V. Selection of endogenous reference genes for qRT-PCR analysis in camelina sativa and identification of FLOWERING LOCUS C allele-specific markers to differentiate summer- and winter-biotypes. *Ind. Crops Prod.* **129**, 495–502 (2019).
56. Bloch, D. et al. Exocyst sect. 3 and phosphoinositides define sites of exocytosis in pollen tube initiation and growth. *Plant. Physiol. Pp*. https://doi.org/10.1104/pp.16.00690 (2016). .00690.2016.
57. Zhang, Y., Immink, R., Liu, C. M., Emons, A. M. & Ketelaar, T. The Arabidopsis exocyst subunit SEC3A is essential for embryo development and accumulates in transient puncta at the plasma membrane. *New. Phytol.* **199**, 74–88 (2013).
58. Zhao, Y. et al. Whole-genome survey and characterization of MADS-box gene family in maize and sorghum. *Plant. Cell. Tissue Organ. Cult. (PCTOC).* **105**, 159–173 (2011).
59. Gao, H. et al. Genome-wide survey of potato MADS-box genes reveals that StMADS1 and StMADS13 are putative downstream targets of tuberigen StSP6A. *BMC Genom.* **19**, 726 (2018).
60. Wang, Y. et al. Genome-Wide analysis of the MADS-Box transcription factor family in solanum lycopersicum. *Int. J. Mol. Sci.* **20**, 2961 (2019).
61. Schilling, S., Kennedy, A., Pan, S., Jermiin, L. S. & Melzer, R. Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol.* **225**, 511–529 (2020).
62. Kagale, S. et al. The emerging biofuel crop camelina sativa retains a highly undifferentiated hexaploid genome structure. *Nat. Commun.* **5**, 3706 (2014).
63. Mun, J. H. et al. Genome-wide comparative analysis of the brassica Rapa gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol.* **10**, R111 (2009).
64. Nardeli, S. M. et al. Genome-wide analysis of the MADS-box gene family in polyploid cotton (Gossypium hirsutum) and in its diploid parental species (Gossypium arboreum and Gossypium raimondii). *Plant Physiol. Biochem.* **127**, 169–184 (2018).
65. Huang, X. C., German, D. A. & Koch, M. A. Temporal patterns of diversification in brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. *Ann. Bot.* **125**, 29–47 (2020).
66. Gramzow, L., Weilandt, L. & Theißen, G. MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Ann. Bot.* **114**, 1407–1429 (2014).
67. Airoldi, C. A. & Davies, B. Gene duplication and the evolution of plant MADS-box transcription factors. *J. Genet. Genomics*. **39**, 157–165 (2012).
68. Erwin, D. H. Evolutionary dynamics of gene regulation. in 407–431 (2020). https://doi.org/10.1016/bs.ctdb.2020.02.006
69. Heidari, P., Ahmadizadeh, M. & Najafi-Zarrini, H. In Silico analysis of Cis-Regulatory elements on Co-Expressed genes. *J. BIOL. ENVIRON. SCI.* **9** (2015). https://www.genevestigator.com/gv
70. De Bodt, S., Theissen, G. & Van De Peer, Y. Promoter analysis of MADS-box genes in eudicots through phylogenetic footprinting. *Mol. Biol. Evol.* **23**, 1293–1303 (2006).
71. Alejandra Mandel, M., Gustafson-Brown, C., Savidge, B. & Yanofsky, M. F. Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. *Nature* **360**, 273–277 (1992).
72. Chen, F., Zhang, X., Liu, X. & Zhang, L. Evolutionary analysis of MIKCc-Type MADS-Box genes in gymnosperms and angiosperms. *Front. Plant. Sci.* **8**, (2017).
73. Coen, E. S. & Meyerowitz, E. M. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**, 31–37 (1991).
74. Searle, I. et al. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. *Genes Dev.* **20**, 898–912 (2006).
75. Sheldon, C. C., Rouse, D. T., Finnegan, E. J., Peacock, W. J. & Dennis, E. S. The molecular basis of vernalization: The central role of FLOWERING LOCUS C (FLC). *Proceedings of the National Academy of Sciences* 97, 3753–3758 (2000).
76. Diaz-Riquelme, J., Lijavetzky, D., Martinez-Zapater, J. M. & Carmona, M. J. Genome-wide analysis of MIKCC-type MADS box genes in grapevine. *Plant. Physiol.* **149**, 354–369 (2009).
77. Walden, A. R., Wang, D. Y., Walter, C. & Gardner, R. C. A large family of TM3 MADS-box cDNAs in Pinus radiata includes two members with deletions of the conserved K domain. *Plant. Sci.* **138**, 167–176 (1998).
78. Winter, K-U. et al. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. U S A.* **96**, 7342–7347 (1999).
79. Folsom, J. J., Begcy, K., Hao, X., Wang, D. & Walia, H. Rice Fertilization-Independent endosperm1 regulates seed size under heat stress by controlling early endosperm development. *Plant. Physiol.* **165**, 238–248 (2014).
80. Pařenicová, L. et al. Molecular and phylogenetic analyses of the complete MADS-Box transcription factor family in arabidopsis: new openings to the MADS world. *Plant. Cell.* **15**, 1538–1551 (2003).
81. de Folter, S., Busscher, J., Colombo, L., Losa, A. & Angenent, G. C. Transcript profiling of transcription factor genes during silique development in Arabidopsis. *Plant. Mol. Biol.* **56**, 351–366 (2004).
82. Portereiko, M. F. et al. AGL80 is required for central cell and endosperm development in Arabidopsis. *Plant. Cell.* **18**, 1862–1872 (2006).
83. Comstock, J. P. Hydraulic and chemical signalling in the control of stomatal conductance and transpiration. *J. Exp. Bot.* **53**, 195–200 (2002).

84. Chaves, M. M. & Oliveira, M. M. Mechanisms underlying plant resilience to water deficits: prospects for water-saving agriculture. *J. Exp. Bot.* **55**, 2365–2384 (2004).
85. Yoshida, T., Mogami, J. & Yamaguchi-Shinozaki, K. ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Curr. Opin. Plant. Biol.* **21**, 133–139 (2014).
86. Huo, S. et al. Genome-wide analysis of the MADS-box gene family in rhododendron Hainanense Merr. And expression analysis under heat And waterlogging stresses. *Ind. Crops Prod.* **172**, (2021).
87. Wei, M., Wang, Y., Pan, R. & Li, W. Genome-Wide identification and characterization of MADS-box family genes related to floral organ development and stress resistance in Hevea Brasiliensis müll. *Arg Forests*. **9**, 304 (2018).

Author contributions

S. T., P.J. and M. M. conceived and design the study. S. T., P.J., M. M., A.M. and P.H. organized and performed the experiments. S. T., P.J. and M. M. were involved in data interpretation. S.T., M.M. and P. H. wrote the manuscript. P. J. and M. M. planned and supervised the study and edited the final version of the manuscript. All authors read and approved the final version of the manuscript.

Funding

Not applicable.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-93724-9.

Correspondence and requests for materials should be addressed to P.J. or P.H.

Reprints and permissions information is available at www.nature.com/reprints.